

Social Phenomena in Global Networks

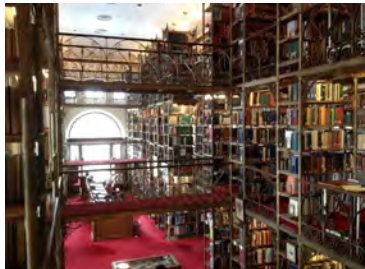
Jon Kleinberg

Cornell University



Including joint work with Lars Backstrom, Cristian Danescu-Niculescu-Mizil,
Dan Huttenlocher, Xiangyang Lan, Lillian Lee, Cameron Marlow,
Bo Pang, and Johan Ugander

The Many Faces of Computing



The Earth, as Seen from a Photo-Sharing Site



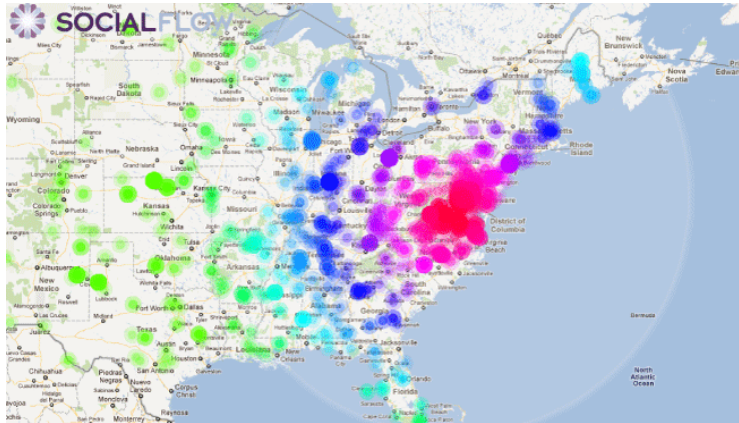
Social Sensing (Hurricane Dean)



Social Sensing (Hurricane Dean)



Social Sensing (Virginia Earthquake)



WHOA!
EARTHQUAKE!



Social Media and Collective Effort

Systems designed to channel effort into collective outcomes.

- reCAPTCHA, foldit, Galaxy Zoo, eBird, DARPA Network Challenge, ...
 - StackOverflow, MathOverflow, Quora, Wikipedia, Polymath, ...
-

Many settings: social media's role is intentional but less structured:

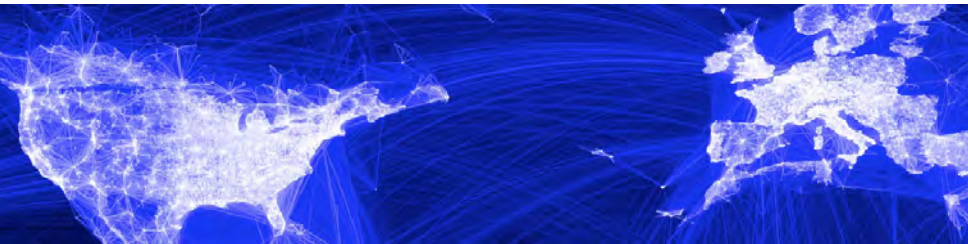
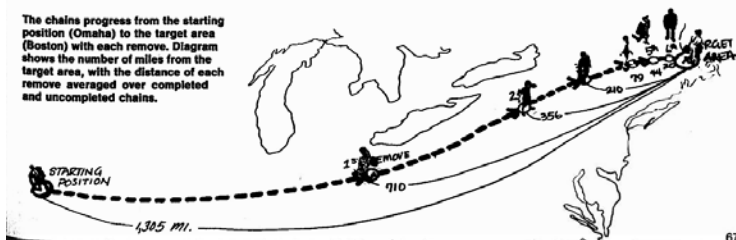


Wu et al 2012



Research Challenges

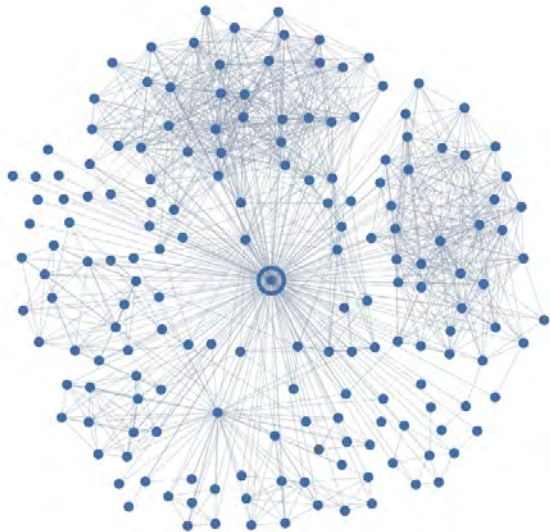
- (1) Designing for massive interacting audiences.
- (2) A radically new way to study questions in the social sciences.



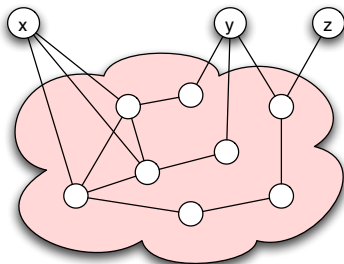
Hard questions in the social sciences:

- What is the probability you form new friendships or engage in new activities based on behaviors of existing friends?
[Backstrom-Huttenlocher-Kleinberg-Lan 2006, Kossinets-Watts 2006, Leskovec-Adamic-Huberman 2006, Romero-Kleinberg-Meeder 2011, Ugander et al 2012]
- Why are you similar to your friends? Because they influence you, or because you seek out people who are already similar?
[Anagnostopoulos-Kumar-Mahdian 2008, Crandall-Cosley-Huttenlocher-Kleinberg-Suri 2008, Aral-Muchnik-Sundararajan 2009, Shalizi-Thomas 2011]
- How do you evaluate and form opinions about other people? How are these opinions reflected in your interactions?
[Danescu-Niculescu-Mizil-Kossinets-Kleinberg-Lee 2009, Leskovec-Huttenlocher-Kleinberg 2010, Szell-Lambiotte-Thurner 2010, Tan-Lee-Tang-Jiang-Zhou-Li 2011]

One person's neighborhood in a social network



Joining Groups

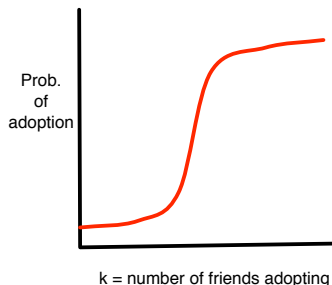
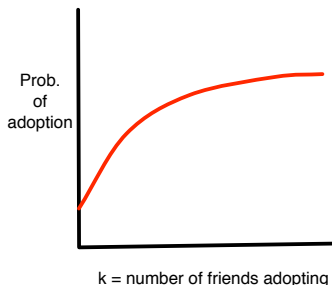


Social networks where people make decisions about new behaviors.

- User-defined groups in on-line communities; affiliation with professional communities; decision to use a hashtag on Twitter; ...
- What factors predict your tendency to adopt?

Diffusion Curves

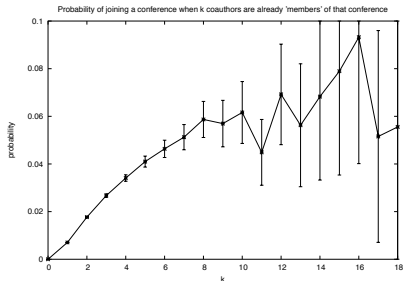
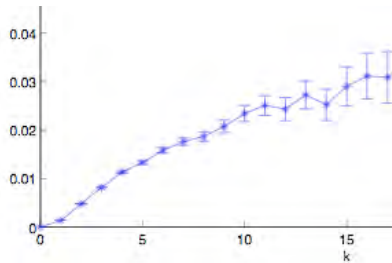
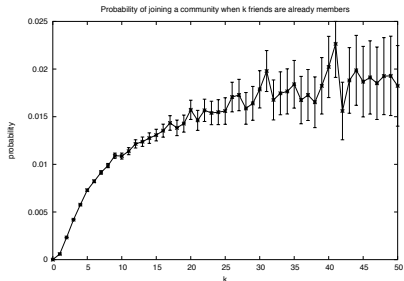
Long-standing framework: probability of adopting a behavior depends on number of network neighbors already adopting.



Key issue: qualitative shape of the curves.

- Diminishing returns? Critical mass?

Diffusion Curves



- (a) Joining a LiveJournal group [Backstrom et al. 06]
- (b) Editing a Wikipedia article [Crandall et al. 08]
- (c) Publishing at a CS conference [Backstrom et al 06]

Prediction and Causality

You're more likely to do something when more friends are doing it. Is that because

- They influenced you to do it?, or
- You're similar to your friends, and subject to similar external factors?

[Cohen 1977, Kandel 1978, Aral et al. 2009, Shalizi-Thomas 2011].



(a)



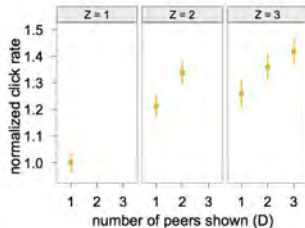
(b)



(c)

An experiment to sort out these effects

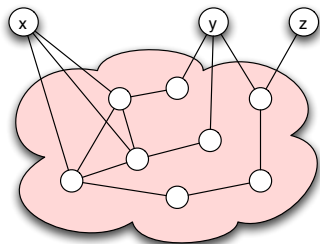
[Bakshy-Eckles-Yan-Rosenn 2012]



Structural Diversity

Dependence on number of friends:
a first step toward general prediction.

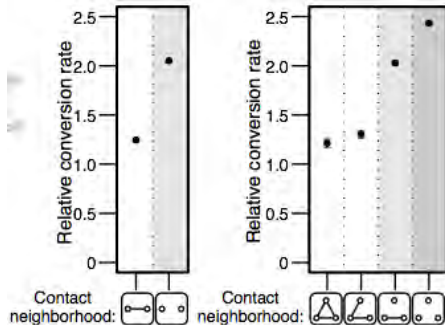
- Given the full pattern of connections among your friends, estimate probability of adopting a new behavior.



Connectedness of friends

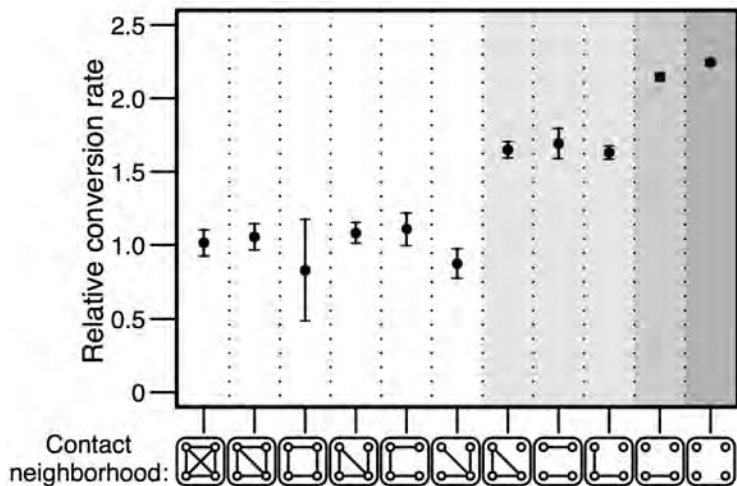
[Ugander-Backstrom-Marlow-Kleinberg]

- Data from invitations to join Facebook.

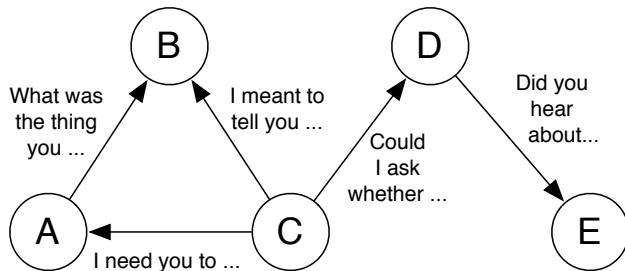


Structural Diversity

With four neighbors:



Status, Power, and Language



Can we detect the influence of people on each other through patterns in their language use?

[Diehl et al 2007, Scholand et al 2010, Gilbert 2012].

Signals that work across domains using language coordination

- Danescu-Niculescu-Mizil-Lee-Pang-Kleinberg 2012.

Language Coordination



Doc: At least you were outside.
Carol: It doesn't make much difference ...

Unconscious imitation of function-word choices (prepositions, articles, conjunctions, quantifiers, ...) [Niederhoffer-Pennebaker 2002].

- **Related to other kinds of coordination:**
posture [Condon-Ogston 1967], pause length [Jaffe-Feldstein 1970], ...
- **Conversations are unconsciously “patterned and coordinated, like a dance.”**

Hypothesis: When there is a status or power imbalance, the low-power person coordinates more and the high-power person coordinates less.

Language Coordination



Doc: At least you were outside.
Carol: It doesn't really matter ...

Unconscious imitation of function-word choices (prepositions, articles, conjunctions, quantifiers, ...) [Niederhoffer-Pennebaker 2002].

- **Related to other kinds of coordination:**
posture [Condon-Ogston 1967], pause length [Jaffe-Feldstein 1970], ...
- **Conversations are unconsciously “patterned and coordinated, like a dance.”**

Hypothesis: When there is a status or power imbalance, the low-power person coordinates more and the high-power person coordinates less.

Language Coordination



Doc: At least you were outside.
Carol: It's not really important ...

Unconscious imitation of function-word choices (prepositions, articles, conjunctions, quantifiers, ...) [Niederhoffer-Pennebaker 2002].

- **Related to other kinds of coordination:**
posture [Condon-Ogston 1967], pause length [Jaffe-Feldstein 1970], ...
- **Conversations are unconsciously “patterned and coordinated, like a dance.”**

Hypothesis: When there is a status or power imbalance, the low-power person coordinates more and the high-power person coordinates less.

Language Coordination



Doc: At least you were outside.
Carol: It doesn't make much difference ...

Unconscious imitation of function-word choices (prepositions, articles, conjunctions, quantifiers, ...) [Niederhoffer-Pennebaker 2002].

- **Related to other kinds of coordination:**
posture [Condon-Ogston 1967], pause length [Jaffe-Feldstein 1970], ...
- **Conversations are unconsciously “patterned and coordinated, like a dance.”**

Hypothesis: When there is a status or power imbalance, the low-power person coordinates more and the high-power person coordinates less.

Measuring Coordination



Doc: At least you were outside.
Carol: It doesn't make much difference ...

A method for measuring the extent of coordination in text data.

[Danescu-Niculescu-Mizil-Dumais-Gamon 2011]

- When A uses a word of type T , does this raise B 's chance of using it in the next utterance?
- Control for the base rate at which B uses T with A in general.

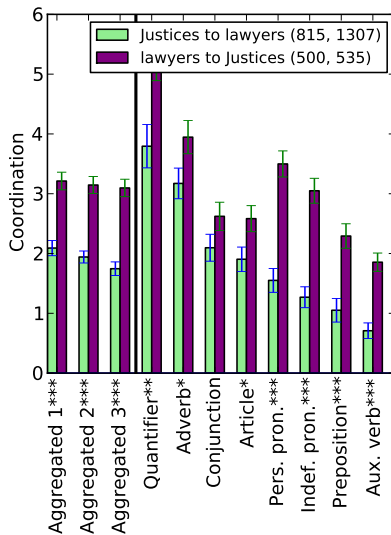
Coordination measure:

- $\Pr[B \text{ uses } T \mid A \text{ used } T \text{ in previous utterance}] - \Pr[B \text{ uses } T \mid B \text{ is replying to } A]$

A First Test: Supreme Court Data



Transcripts of oral arguments from approximately 200 Supreme Court cases.



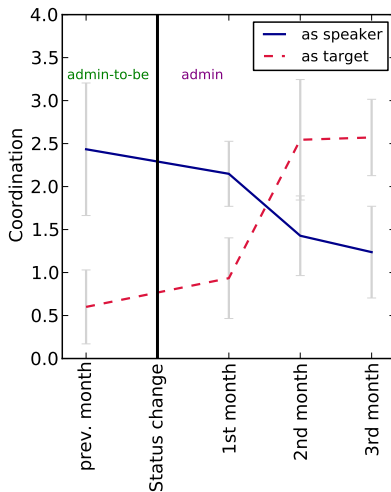
Change over Time



WIKIPEDIA
The Free Encyclopedia

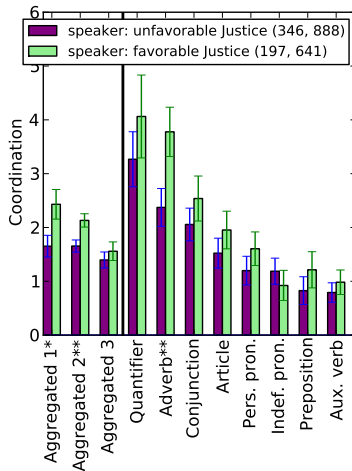
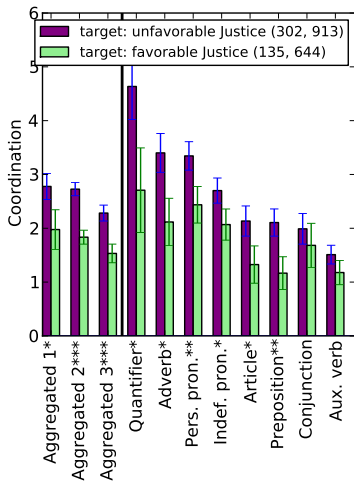
Conversations on Wikipedia talk pages.

What if a person experiences a change in status?



Dependence and Power

Power differences can also arise from dependence [Emerson 1962].



- Lawyers coordinate more to Justices who end up voting against them.
- Justices coordinate more to lawyers they end up voting in favor of.

Final Reflections: Privacy and Data-Sharing

Many research datasets arise from settings where users have expectations of privacy.

- How do we help users reason about consequences of public sharing?
- To what extent can organizations share research datasets?

Challenge: many indirect channels for information to be revealed.

- Times/places at which you did certain actions often uniquely identifying
Narayanan-Shmatikov 2006.
- The pattern of links among your friends is often uniquely identifying
Backstrom-Dwork-Kleinberg 2007.
- Your friends' characteristics can form the basis for predictions about you
Zheleva-Getoor 2009, Jernigan-Mistree 2009, Mislove et al 2010.
- If two people do similar things at similar times, likely know each other
Crandall et al 2010.

Approaches to mitigate these risks in data-sharing via differential privacy [Dwork 2006], census procedures [Machanavajjhala et al 2008].

Final Reflections: The Challenge of Prediction

With accurate models and enough data, can we predict outcomes?



MusicLab [Salganik, Dodds, and Watts 2006]:

Music download site: preview songs, then download.

A “leaderboard” showing the most downloaded songs.

Can we predict which songs will come out on top?

Final Reflections

MySpace is doubly awkward because it makes public what should be private. It doesn't just create social networks, it anatomizes them. It spreads them out like a digestive tract on the autopsy table. You can see what's connected to what, who's connected to whom.

– Toronto Globe and Mail, June 2006.



- Social networks — implicit for millenia — are being recorded at high resolution.
- The opportunity: to understand why social processes turn out the way they do, with insights for design, analysis, prediction.
- Computational ideas will be crucial in all these developments.